

## ESTIMATING THE MEAN OF A NON-LINEAR FUNCTION OF CONDITIONAL EXPECTATION

L. Jeff Hong

Department of Industrial Engineering and Logistics Management  
Hong Kong University of Science and Technology  
HONG KONG

Sandeep Juneja

School of Technology and Computer Science  
Tata Institute of Fundamental Research  
Mumbai, 400005, INDIA

### ABSTRACT

Consider the problem of estimating the expectation of a non linear function of a conditional expectation. This function is allowed to be non-differentiable and discontinuous at a finite set of points to capture practical settings. We develop a nested simulation strategy to estimate this via simulation and identify bias and optimized mean square error allocation. We show that this mean square error converges to zero at the rate  $\Gamma^{-2/3}$ , as  $\Gamma \rightarrow \infty$ , where  $\Gamma$  denotes the available computational budget. We also consider combining nested simulation technique with kernel based estimation methods. We note that while the kernel based method have a better convergence rate when the underlying random process has dimensionality less than or equal to three, pure nested simulation may be preferred when this dimension is above four.

### 1 INTRODUCTION

In this paper we consider the problem of estimating via Monte Carlo, the quantity

$$\alpha = E(G(EX|Y))$$

where  $X$  is a random variable in  $\mathfrak{R}$ ,  $Y$  is a random vector in  $\mathfrak{R}^d$ ,  $d \geq 1$ , and  $G$  is a non-linear function that may have finite number of discontinuous/non-differentiable points. This has application in many practical settings particularly in pricing and risk management in finance. Examples include pricing compound options (see, for instance, Glasserman 2004). As an example, consider a call option expiring at time  $T_1$  to buy a call option expiring at time  $T_2 > T_1$ . Assuming for notational simplicity that the discount rate is zero, the price of this compound option may be expressed as

$$E \left[ E(S(T_2) - K_2)^+ | S(T_1) - K_1 \right]^+.$$

where  $S(t)$  denotes the price of the underlying stock at time  $t$ ,  $K_2$  denotes the strike price of a call option on this stock that expires at time  $T_2$ , and  $K_1$  denotes the strike price of an option on this call option that expires at time  $T_1$ .

Similarly, a price of a CDO tranche has the form  $\alpha$ . To see this, note that the discounted loss amount in the basket of credit instruments in a CDO may be expressed as  $E(X|Y)$  where  $Y$  denotes the underlying random variables such as realization of credit intensities, default event indicators, etc. at a given time point. Conditioned on these, the value of each instrument is its risk neutral price and hence may be a conditional expectation. This then makes the total discounted loss amount a conditional expectation (see, for instance Gordy and Juneja 2006). Owner of a tranche of a CDO with an attachment point  $a$  and detachment point  $b$  will suffer losses when the overall loss exceeds  $a$ , with the loss amount capped once the overall loss exceeds  $b$ . Precisely, the loss suffered equals

$$\min(b, \max(E(X|Y), a)) - a.$$

In risk management of a portfolio comprising sophisticated derivatives one encounters the problem of estimating  $\alpha$  (see, for example, Lee 1998, Gordy and Juneja 2008) where  $G(x)$ , for instance, may be of the form  $(x - u)^+$  for a positive  $u$ ,  $I(x > u)$  for an indicator function  $I(\cdot)$ , or  $x^n$  for some  $n > 0$ .

One approach to estimating  $\alpha$  is via nested simulation: Many independent samples of  $Y$  are generated in the ‘outer loop’ and for each such  $Y$  many independent samples of  $X$  are generated in the ‘inner loop’ to estimate the associated  $EX|Y$ . We analyze the bias of the resultant estimator. In the spirit of the analysis conducted by Lee 1998 and Gordy and Juneja 2008 in a more specialized settings, we also discuss the optimal computational resource allocation between the inner and outer loop to minimize the resultant mean square error of the estimator. In particular, we see that even in this general setting, the mean square decays at the rate  $\Gamma^{-2/3}$ , as  $\Gamma \rightarrow \infty$ , where  $\Gamma$  denotes the available computational budget.

We also introduce a combination of the kernel methods and nested simulation based estimator for  $\alpha$  (see, for instance, Bosq 1998, Li and Racine 2007, and Pagan and Ullah 1999 for recent overviews on kernel methods). To keep the analysis simple, here we assume that  $G(\cdot)$  is smooth everywhere. We show that the optimal mean square error decays at rate

$$O(\Gamma^{-\min(1, \frac{4}{d+2})}).$$

Hence, when the dimensionality of  $Y$  is less than or equal to two, the mean square error has decay rate  $O(\Gamma^{-1})$ . When it is three, the decay rate  $O(\Gamma^{-4/5})$  is still an improvement over pure nested simulation. However, for  $d \geq 5$ , the pure nested simulation may be preferred. We make simplifying assumptions and use heuristical arguments in this paper to aid the exposition. A more rigorous expanded version will appear separately.

In Section 2 we discuss nested simulation estimation. Its combination with kernel based methods is discussed in Section 3. We end with a brief conclusion in Section 4. Some of the proof outlines are relegated to the appendix.

## 2 NESTED SIMULATION ESTIMATION

Suppose that random vector  $Y$  take values in  $\mathfrak{R}^d$ ,  $d \geq 1$ ,  $X$  takes values in  $\mathfrak{R}$  and  $G: \mathfrak{R} \rightarrow \mathfrak{R}$ . Consider the following procedure to estimate  $\alpha$ : Generate i.i.d. samples  $Y_1, \dots, Y_n$  of  $Y$ . Conditioned on  $Y_i$ , generate samples  $(X_{ij}(Y_i): j = 1, \dots, m)$  of  $X$ . Then

$$\hat{\alpha}_{m,n} = \frac{1}{n} \sum_{i=1}^n G \left( \frac{1}{m} \sum_{j=1}^m X_{ij}(Y_i) \right)$$

is a nested estimator of  $\alpha$ . We first estimate the bias of the estimator  $\hat{\alpha}_{m,n}$ , that is, we analyze

$$|EG(\bar{X}_m(Y)) - \alpha|$$

where  $\bar{X}_m(Y) = \frac{1}{m} \sum_{j=1}^m X_j(Y)$  and  $X_j(Y)$  is distributed as  $X$  conditional on  $Y$ .

In this paper we restrict ourselves to settings where random variable  $H = EX|Y$  has a probability density function  $f_H(\cdot)$ . We use the notation  $H$  with  $EX|Y$  interchangeably, using the former to avoid notational clutter, and the latter where greater clarity is useful. Conditioned on  $Y$ , let  $\sigma^2(Y)$  denote the variance of  $X$ . Under mild conditions, due to the central limit theorem, conditioned on  $Y$

$$\frac{1}{\sqrt{m}} \sum_{j=1}^m X_j(Y)$$

converges to a Gaussian random variable with mean  $EX|Y$  and variance  $\sigma^2(Y)$ . To keep the exposition simple we will assume that conditional on  $Y$ ,  $\frac{1}{\sqrt{m}} \sum_{j=1}^m X_j(Y)$  has a Gaussian distribution (with mean  $EX|Y$  and variance  $\sigma^2(Y)$ ). In more general settings, the deviations from Gaussian distribution are easy to handle using Berry-Esseen inequalities (see Lee 1998 for similar analysis).

Practically important functions  $G$  may have a finite number of points where they are non-differentiable or even discontinuous. To understand the bias for such  $G$ , it suffices to consider functions  $G$  that are truncated at a point. For instance, either  $G(x) = 0$  for  $x \geq b$  or  $G(x) = 0$  for  $x \leq a$  (See, Figure 1). Once, we have established the bias for both such functions, the bias for a general function with finite non-differentiability or even discontinuity points, can be evaluated by adding or subtracting such functions for different values of  $a$  and  $b$ .

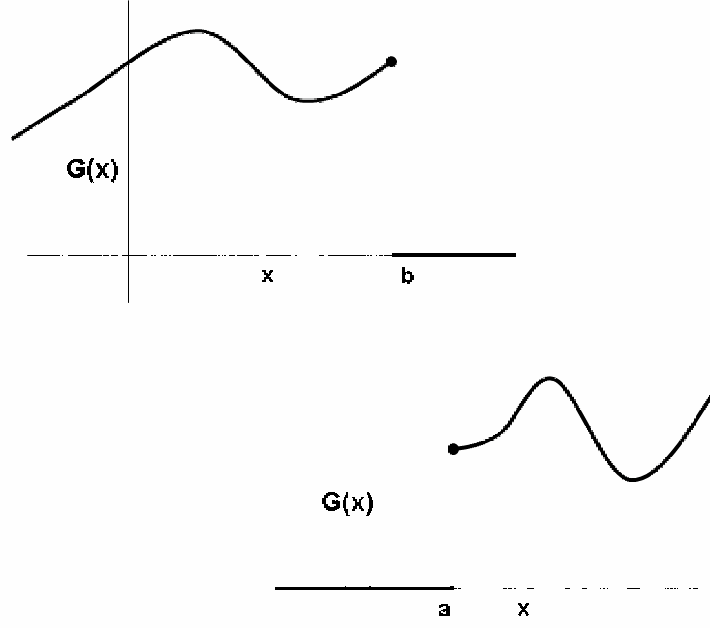


Figure 1: Practically important functions  $G$  may have a finite number of points where they are non-differentiable or even discontinuous. To understand the bias for such  $G$ , it suffices to consider truncated functions shown above. The bias for general functions, when  $EX|Y$  is a continuous random variable, can be evaluated by adding or subtracting such functions for different values of  $a$  and  $b$ .

### 2.1 Bias Estimation

We first assume that  $G(x) = 0$  for  $x > b$ . For  $x \leq b$ , we assume that  $G(\cdot)$  is thrice continuously differentiable (at  $b$  we only consider the limits from the left) with uniformly bounded third derivative. We show that

$$EG(\bar{X}_m(Y)) - \alpha = \frac{1}{2m} EG''(EX|Y)\sigma^2(Y)I(EX|Y \leq b) - \frac{G'(b)f_H(b)}{2m} E(\sigma^2(Y)|EX|Y = b) + \frac{G(b)}{2m} \frac{d}{db} f_H(b) E(\sigma^2(Y)|EX|Y = b) \quad (1)$$

plus smaller order terms, where  $G'(\cdot)$  and  $G''(\cdot)$  denote the first and the second derivatives, respectively.

One can similarly show that if  $G(x) = 0$  for  $x < a$ , and for  $x \geq a$ ,  $G(\cdot)$  is thrice continuously differentiable (at  $a$  we only consider the limits from the right) with uniformly bounded third derivative, then

$$EG(\bar{X}_m(Y)) - \alpha = \frac{1}{2m} EG''(EX|Y)\sigma^2(Y)I(EX|Y \geq a) + \frac{G'(a)f_H(a)}{2m} E(\sigma^2(Y)|EX|Y = a) - \frac{G(a)}{2m} \frac{d}{da} f_H(a) E(\sigma^2(Y)|EX|Y = a) \quad (2)$$

plus smaller order terms.

**Example 1.** Suppose that  $G(\cdot)$  is thrice continuously differentiable everywhere on the real line, with uniformly bounded third derivative, then it follows from above that

$$EG(\bar{X}_m(Y)) - \alpha = \frac{1}{2m} E(G''(EX|Y)\sigma^2(Y)) + \text{smaller order terms.}$$

Hence,

$$E(\bar{X}_m(Y))^2 - E(EX|Y)^2 = \frac{1}{2m}E(\sigma^2(Y)) + \text{smaller order terms.}$$

**Example 2.** Suppose that  $G(\cdot)$  is thrice continuously differentiable on  $[a, b]$  and is zero outside this interval. Again, the limits associated with the first and second derivatives at points  $a$  and  $b$  are defined as limits from the right and the limits from the left, respectively. Then, by expressing this function as a difference of two functions, one on  $(-\infty, b)$  and another on  $(-\infty, a)$ , both sufficiently smooth in their domains, we get

$$\begin{aligned} EG(\bar{X}_m(Y)) - \alpha &= \frac{1}{2m}EG''(EX|Y)\sigma^2(Y)I(a \leq EX|Y \leq b) - \frac{G'(b)f_H(b)}{2m}E(\sigma^2(Y)|EX|Y = b) \\ &+ \frac{G(b)}{2m} \frac{d}{db} f_H(b)E(\sigma^2(Y)|EX|Y = b) + \frac{G'(a)f_H(a)}{2m}E(\sigma^2(Y)|EX|Y = a) - \frac{G(a)}{2m} \frac{d}{db} f_H(a)E(\sigma^2(Y)|EX|Y = a) + \text{smaller order terms.} \end{aligned}$$

**Example 3.** Consider  $G(x) = I(x \geq u)$  for a constant  $u$ . Then,

$$P(\bar{X}_m(Y) \geq u) - P(EX|Y \geq u) = -\frac{1}{2m} \frac{d}{db} f_H(u)E(\sigma^2(Y)|EX|Y = u) + \text{smaller order terms.}$$

This was reported in Lee 1998 and Gordy and Juneja 2008.

**Example 4.** Consider  $G(x) = (x - u)^+$  for a constant  $u$ . Then,

$$E(\bar{X}_m(Y) - u)^+ - E(EX|Y - u)^+ = +\frac{f_H(u)}{2m}E(\sigma^2(Y)|EX|Y = u) + \text{smaller order terms.}$$

**Example 5.** Consider

$$G(x) = \min(b, \max(x, a)) - a = (x - a)^+ - (x - b)^+,$$

for  $a < b$ . Then,

$$EG(\bar{X}_m(Y)) - \alpha = \frac{f_H(a)}{2m}E(\sigma^2(Y)|EX|Y = a) - \frac{f_H(b)}{2m}E(\sigma^2(Y)|EX|Y = b) + \text{smaller order terms.}$$

**Proof outline :** To see (1), consider the following decomposition:

$$E(G(\bar{X}_m(Y)) - G(EX|Y)I(EX|Y \leq b, \bar{X}_m(Y) \leq b)), \tag{3}$$

$$E(G(\bar{X}_m(Y)) - G(EX|Y)I(EX|Y \leq b, \bar{X}_m(Y) \geq b)), \tag{4}$$

and

$$E(G(\bar{X}_m(Y)) - G(EX|Y)I(EX|Y > b, \bar{X}_m(Y) \leq b)). \tag{5}$$

We show that (3) equals

$$-\frac{1}{2m}G'(b)f_H(b)E(\sigma^2(Y)|H = b) + \frac{1}{2m}E\sigma^2(Y)G''(EX|Y)I(EX|Y \leq b) + \text{smaller order terms.} \tag{6}$$

Furthermore, (4) equals

$$-\frac{G(b)f_H(b)}{\sqrt{m}}E(WI(W \geq 0)|H = b) + \frac{G(b)}{4m} \frac{df_H(b)}{db} E(\sigma^2(Y)|H = b) + \frac{G'(b)f_H(b)}{4m}E(\sigma^2(Y)|H = b) \tag{7}$$

plus smaller order terms, and (5) equals

$$-\frac{G(b)f_H(b)}{\sqrt{m}}E(WI(W \leq 0)|H = b) + \frac{G(b)}{4m} \frac{df_H(b)}{db} E(\sigma^2(Y)|H = b) - \frac{G'(b)f_H(b)}{4m} E(\sigma^2(Y)|H = b) \quad (8)$$

plus smaller order terms. Then, noting that

$$E(W|H = b) = E(E(W|Y)|H = b) = 0,$$

(1) follows. The fact that equations (3), (4) and (5) equal (6), (7) and (8), respectively is shown in the appendix.

## 2.2 Optimizing Mean Square Error

Note that the mean square error of  $\hat{\alpha}_{m,n}$  equals its squared bias plus its variance. As we discussed above, the squared bias has the expression  $c^2/m^2 + o(1/m^2)$  for an appropriate constant  $c$ . We now evaluate the dominant terms of the variance of  $\hat{\alpha}_{m,n}$ . For any rv  $X$ , let  $V(X)$  denote its variance. Then,

$$V(\hat{\alpha}_{m,n}) = \frac{1}{n} V(G(\bar{X}_m(Y))),$$

and

$$V(G(\bar{X}_m(Y))) = EG(\bar{X}_m(Y))^2 - (EG(\bar{X}_m(Y)))^2.$$

From the above analysis, it is clear that under mild conditions,

$$EG(\bar{X}_m(Y))^2 = EG(EX|Y)^2 + O(1/m),$$

and

$$(EG(\bar{X}_m(Y)))^2 = (EG(EX|Y))^2 + O(1/m^2).$$

Therefore,

$$V(\hat{\alpha}_{m,n}) = \frac{1}{n} V(G(EX|Y)) + \frac{1}{n} O(1/m).$$

As in Lee 1998 and Gordy and Juneja 2008, we now consider the problem of finding asymptotically optimal  $m$  and  $n$ , when the overall computational budget  $\Gamma$  increases to  $\infty$ . If we assume that the average computational effort in the outer loop is a constant  $\gamma$  and the average computational effort in the inner loop is  $m\beta$ , then for large values of  $n$ , the overall computational effort is close to  $n(\gamma + m\beta)$ . We then look for  $m$  and  $n$  that minimize the mean square error when the total computational budget  $\Gamma = n(\gamma + m\beta)$  and  $\Gamma \rightarrow \infty$ .

Note that the dominant terms in the mean square error are

$$\frac{c^2}{m^2} + \frac{v}{n} \quad (9)$$

where  $v = V(G(EX|Y))$ . Setting  $n = \Gamma/(\gamma + m\beta)$  and optimizing (9) over  $m$ , we see that the optimal inner loops

$$m^* = \left( \frac{2c}{v\beta} \right)^{1/3} \Gamma^{1/3}.$$

Then optimal outer loops

$$n^* = \frac{\Gamma}{v + \beta m^*} = \left( \frac{v\beta}{2c} \right)^{1/3} \Gamma^{2/3} + o(\Gamma^{2/3}).$$

The mean square error has a decay rate of order  $\Gamma^{-2/3}$ .

### 3 KERNEL METHOD BASED ESTIMATION

Recall that in the nested simulation method for estimating  $\alpha$ , we generate independent samples of  $Y$ , ( $Y_i : i \leq n$ ) and for each  $i$ , we generate independent samples of  $X$ , ( $X_j : j \leq m$ ), conditioned on  $Y_i$ . The estimator of each  $EX|Y_i$  is  $\bar{X}_m(Y_i)$  and the estimator of  $\alpha$  is  $\frac{1}{n} \sum_{i=1}^n G(\bar{X}_m(Y_i))$ . Another way suggested by kernel based methods is to also include information associated with  $Y_j$ ,  $j \neq i$  for estimating  $EX|Y_i$ , giving higher weights to  $Y_j$ 's that are in the vicinity of  $Y_i$ .

Specifically, consider a kernel function  $K : \mathfrak{R} \rightarrow \mathfrak{R}^+$ . Typically such kernels are assumed to be bounded and symmetric such that  $yK(y) \rightarrow 0$  as  $\|y\| \rightarrow \infty$  and  $\int_{\mathfrak{R}} y^2 K(y) dy < \infty$ . Then an estimator for  $EX|Y_i$  is given as

$$\frac{\sum_{j=1}^n \bar{X}_m(Y_j) K\left(\frac{\|Y_j - Y_i\|}{\delta_n}\right)}{\sum_{j=1}^n K\left(\frac{\|Y_j - Y_i\|}{\delta_n}\right)}$$

where  $\{\delta_n\}$  is a sequence that decreases to zero and  $n\delta_n^d \rightarrow \infty$  (recall that  $d$  is the dimension of  $Y$ ). In this paper, we focus on a simple kernel  $K(x) = I(x < 1)$ .

First consider the case, where  $Y$  takes values in  $\mathfrak{R}$ . Then our estimator of  $EX|Y_i$  is

$$\tilde{X}_m(Y_i) = \frac{\sum_{j=1}^n \bar{X}_m(Y_j) I(|Y_i - Y_j| < \delta_n)}{\sum_{j=1}^n I(|Y_i - Y_j| < \delta_n)}.$$

Our estimator of  $\alpha$  is

$$\tilde{\alpha}_{m,n} = \frac{1}{n} \sum_{i=1}^n G(\tilde{X}_m(Y_i)).$$

In this section we assume that  $G(\cdot)$  is sufficiently smooth everywhere in  $\mathfrak{R}$ .

#### 3.1 Bias Estimation

We first analyze the bias  $|\beta_{m,n}|$  of this estimator, where

$$\beta_{m,n} = E(\tilde{\alpha}_{m,n}) - \alpha.$$

Denote any vector ( $Z_i : i \leq n$ ) by  $\vec{Z}_n$ . Let  $\eta(y) = EX|Y = y$ . We assume that this function is twice continuously differentiable. Note that

$$\tilde{X}_m(Y_i) = Z_{m,n} + \frac{\sum_{j=1}^n \eta(Y_j) I(|Y_i - Y_j| < \delta_n)}{\sum_{j=1}^n I(|Y_i - Y_j| < \delta_n)}$$

where

$$Z_{m,n} = \frac{(\sum_{j=1}^n \bar{X}_m(Y_j) - \eta(Y_i)) I(|Y_i - Y_j| < \delta_n)}{\sum_{j=1}^n I(|Y_i - Y_j| < \delta_n)}.$$

Furthermore,

$$\frac{\sum_{j=1}^n \eta(Y_j) I(|Y_i - Y_j| < \delta_n)}{\sum_{j=1}^n I(|Y_i - Y_j| < \delta_n)} = \eta(Y_i) + R_1(\vec{Y}_n, \delta_n) + R_2(\vec{Y}_n, \delta_n)$$

where,

$$R_1(\vec{Y}_n, \delta_n) = \eta'(Y_i) \frac{\sum_{j=1}^n (Y_j - Y_i) I(|Y_i - Y_j| < \delta_n)}{\sum_{j=1}^n I(|Y_i - Y_j| < \delta_n)},$$

and

$$R_2(\vec{Y}_n, \delta_n) = \frac{\eta''(Y_i) \sum_{j=1}^n (Y_j - Y_i)^2 I(|Y_i - Y_j| < \delta_n)}{2 \sum_{j=1}^n I(|Y_i - Y_j| < \delta_n)} + \text{remainder term.}$$

Hence,

$$\tilde{X}_m(Y_i) = Z_{m,n} + \eta(Y_i) + R_1(\vec{Y}_n, \delta_n) + R_2(\vec{Y}_n, \delta_n).$$

Note that  $\beta_{m,n}$  equals the expectation of

$$G'(\eta(Y_i))(Z_{m,n} + R_1(\vec{Y}_n, \delta_n) + R_2(\vec{Y}_n, \delta_n)) + \frac{G''(\eta(Y_i))}{2} (Z_{m,n} + R_1(\vec{Y}_n, \delta_n) + R_2(\vec{Y}_n, \delta_n))^2 + \text{remainder term.}$$

Since,

$$E(Z_{m,n} | \vec{Y}_n) = 0$$

it follows that  $\beta_{m,n}$  equals

$$E\left(G'(\eta(Y_i))(R_1(\vec{Y}_n, \delta_n) + R_2(\vec{Y}_n, \delta_n))\right) + \frac{1}{2} E Z_{m,n}^2 G''(\eta(Y_i)) + \text{remainder terms.}$$

We now evaluate the dominant terms in  $E(G'(\eta(Y_i))(R_1(\vec{Y}_n, \delta_n)))$ ,  $E(G'(\eta(Y_i))R_2(\vec{Y}_n, \delta_n))$  and  $E Z_{m,n}^2 G''(\eta(Y_i))$ .

For this, the following results are useful and are easily established. Let  $f_Y(\cdot)$  denote the pdf of  $Y$  which we assume is sufficiently smooth. Then,

$$P(|Y_j - y| < \delta_n) = 2\delta_n f_Y(y) + O(\delta_n^2),$$

$$E(Y_j - y)I(|Y_j - y| < \delta_n) = \frac{2}{3} \delta_n^3 f_Y'(y) + o(\delta_n^3),$$

and

$$E(Y_j - y)^2 I(|Y_j - y| < \delta_n) = \frac{2}{3} \delta_n^3 f_Y(y) + o(\delta_n^3).$$

Conditioning on  $Y_i$ , using a variant of law of large numbers it can be seen that

$$\frac{1}{n\delta_n} \sum_{j=1}^n I(|Y_i - Y_j| < \delta_n) \rightarrow 2f_Y(Y_i) \text{ a.s.,}$$

$$\frac{1}{n\delta_n^3} \sum_{j=1}^n (Y_j - Y_i)I(|Y_i - Y_j| < \delta_n) \rightarrow \frac{2}{3} f_Y'(Y_i) \text{ a.s.,}$$

and

$$\frac{1}{n\delta_n^3} \sum_{j=1}^n (Y_j - Y_i)^2 I(|Y_i - Y_j| < \delta_n) \rightarrow \frac{2}{3} f_Y(Y_i) \text{ a.s.}$$

It then follows that

$$\frac{1}{\delta_n^2} G'(\eta(Y_i))(R_1(\vec{Y}_n, \delta_n)) \rightarrow \frac{1}{3} G'(\eta(Y_i)) \eta'(Y_i) \frac{f_Y'(Y_i)}{f_Y(Y_i)} \text{ a.s.}$$

Under technical conditions, the LHS is uniformly integrable so that

$$E(G'(\eta(Y_i))(R_1(\vec{Y}_n, \delta_n))) = \frac{\delta_n^2}{3} E G'(\eta(Y_i)) \eta'(Y_i) \frac{f_Y'(Y_i)}{f_Y(Y_i)} + o(\delta_n^2).$$

Similarly,  $E(G'(\eta(Y_i))R_2(\vec{Y}_n, \delta_n))$  can be evaluated. To see this note that

$$\frac{1}{\delta_n^2} G'(\eta(Y_i))(R_2(\vec{Y}_n, \delta_n)) \rightarrow \frac{1}{6} G'(\eta(Y_i))\eta''(Y_i) \text{ a.s.}$$

Therefore, if the LHS is uniformly integrable then,

$$E(G'(\eta(Y_i))R_2(\vec{Y}_n, \delta_n)) = \frac{\delta_n^2}{6} E(G'(\eta(Y_i))\eta''(Y_i)) + o(\delta_n^2).$$

Now consider  $EZ_{m,n}^2 G''(\eta(Y_i))$ . Note that this equals

$$E \left( G''(\eta(Y_i)) \frac{\sum_{j=1}^n (\bar{X}_m(Y_j) - \eta(Y_j))^2 I(|Y_i - Y_j| < \delta_n)}{(\sum_{j=1}^n I(|Y_i - Y_j| < \delta_n))^2} \right).$$

Recall that  $\sigma^2(Y)$  denotes the conditional variance of  $X$  given  $Y$ . Then, above equation equals

$$E \left( \frac{1}{m} G''(\eta(Y_i)) \frac{\sum_{j=1}^n \sigma^2(Y_j) I(|Y_i - Y_j| < \delta_n)}{(\sum_{j=1}^n I(|Y_i - Y_j| < \delta_n))^2} \right). \quad (10)$$

This term can be seen to have a dominant term

$$E \left( \frac{1}{m} G''(\eta(Y_i)) \sigma^2(Y_i) \frac{\sum_{j=1}^n I(|Y_i - Y_j| < \delta_n)}{(\sum_{j=1}^n I(|Y_i - Y_j| < \delta_n))^2} \right).$$

As in our earlier analysis, it follows that under technical conditions, (10) equals

$$\frac{1}{2mn\delta_n} E \left( \frac{G''(\eta(Y_i)) \sigma^2(Y_i)}{f_Y(Y_i)} \right) (1 + o(1)).$$

Therefore, when  $Y$  takes values in  $\mathfrak{X}$ , under technical conditions we have that the bias is dominated by terms of the form

$$d_1 \delta_n^2 + d_2 \frac{1}{mn\delta_n}$$

for constants  $d_1$  and  $d_2$ .

### 3.2 Variance Estimation

We now consider the variance

$$V\left(\frac{1}{n} \sum_{i=1}^n G(\tilde{X}_m(Y_i))\right) \quad (11)$$

and argue that the dominant term has the form  $\frac{v_1}{n}$ . Note that this variance equals

$$\frac{1}{n} V(G(\tilde{X}_m(Y_1))) + (1 - 1/n) \text{Cov}(G(\tilde{X}_m(Y_1)), G(\tilde{X}_m(Y_2)))$$

where  $\text{Cov}(A, B)$  denotes the covariance between random variables  $A$  and  $B$ .

Note that  $V(G(\tilde{X}_m(Y_1)))$  equals

$$EG(\tilde{X}_m(Y_1))^2 - (EG(\tilde{X}_m(Y_1)))^2.$$



From our previous analysis, under technical conditions, this converges to  $V(\eta(Y))$  as  $n \rightarrow \infty$ .

Now consider  $Cov(G(\tilde{X}_m(Y_1)), G(\tilde{X}_m(Y_2)))$ . To ease the analysis we discuss the case where  $G(x) = x$ . In that setting this term equals

$$Cov\left(\frac{\sum_{i=1}^n \eta(Y_i)I(|Y_i - Y_1| < \delta_n)}{\sum_{i=1}^n I(|Y_i - Y_1| < \delta_n)}, \frac{\sum_{j=1}^n \eta(Y_j)I(|Y_j - Y_2| < \delta_n)}{\sum_{j=1}^n I(|Y_j - Y_2| < \delta_n)}\right). \quad (12)$$

Since,

$$\sum_{i=1}^n I(|Y_i - Y_1| < \delta_n) \sim 2n\delta_n f_Y(Y_1)$$

We can show that (12) is asymptotically similar to

$$\frac{1}{(n\delta_n)^2} \sum_{i=1}^n \sum_{j=1}^n Cov\left(\frac{\eta(Y_i)}{2f_Y(Y_1)}I(|Y_i - Y_1| < \delta_n), \frac{\eta(Y_j)}{2f_Y(Y_2)}I(|Y_j - Y_2| < \delta_n)\right). \quad (13)$$

Note that

$$Cov\left(\frac{\eta(Y_i)}{2f_Y(Y_1)}I(|Y_i - Y_1| < \delta_n), \frac{\eta(Y_j)}{2f_Y(Y_2)}I(|Y_j - Y_2| < \delta_n)\right)$$

equals zero when  $i \neq j \neq 1$  or  $2$ . When,  $i = j \neq 1$  or  $2$ , this equals  $O(\delta_n^2)$ . When  $i = 1$  and  $j \geq 3$ , this term is zero. Similarly, when  $j = 2$  and  $i \geq 3$ , this term is zero. When,  $j = 1$  and  $i \geq 3$ , this term is  $O(\delta_n^2)$ . Similarly, when,  $i = 2$  and  $j \geq 3$ , this term is  $O(\delta_n^2)$ . The remaining four terms are bounded from above by  $O(1)$  terms. Then, its easy to see that (13) is dominated by  $O(\frac{1}{n})$ .

This observation generalizes and one can show that (11) has the form  $\frac{v_1}{n}$  plus smaller order terms under technical conditions.

### 3.3 Multi-Dimensional Analysis

We briefly consider the case where  $d > 1$ . The major change in the bias analysis from the case  $d = 1$  is that now

$$R_1(\vec{Y}_n, \delta_n) = \frac{(\sum_{j=1}^n \nabla \eta(Y_i)^T (Y_j - Y_i)I(\|Y_i - Y_j\| < \delta_n))}{\sum_{j=1}^n I(\|Y_i - Y_j\| < \delta_n)},$$

where  $\nabla \eta(y)$  denotes the gradient of  $\eta$  at  $y$  and  $\|y\|$  for  $y \in \mathfrak{R}^d$  denotes its Euclidean norm. Similarly,

$$R_2(\vec{Y}_n, \delta_n) = \frac{1}{2} \frac{(\sum_{j=1}^n (Y_j - Y_i)^T \mathcal{H}(Y_i) (Y_j - Y_i)I(\|Y_i - Y_j\| < \delta_n))}{\sum_{j=1}^n I(\|Y_i - Y_j\| < \delta_n)} + \text{remainder term},$$

where  $\mathcal{H}(y)$  denotes the Hessian of  $\eta$  at  $y$ .

In this case it is easy to see that

$$P(\|Y_j - y\| < \delta_n) = \text{constant} \times \delta_n^d f_Y(y) + o(\delta_n^d),$$

$$E(\nabla \eta(y)^T (Y_j - y)I(\|Y_j - y\| < \delta_n)) = \text{constant} \times \delta_n^{d+2} \nabla \eta(y)^T \nabla f_Y(y) + o(\delta_n^{d+2}),$$

and

$$E(Y_j - y)^T \mathcal{H}(y) (Y_j - y)I(\|Y_j - y\| < \delta_n) = \text{constant} \times \delta_n^{d+2} Tr(\mathcal{H}(y)) f_Y(y) + o(\delta_n^{d+2})$$

where  $Tr(A)$  denotes the trace of matrix  $A$ .

Then, essentially repeating the discussion for  $d = 1$  case, we see that the bias has the dominant terms of the form

$$a\delta_n^2 + \frac{b}{m\delta_n^d}$$

for constants  $a$  and  $b$ .

The discussion on the variance gives a similar conclusion that the dominant term equals  $\frac{v}{n}$  for a constant  $v$ . Then, the mean square error has dominant terms

$$a^2\delta_n^4 + 2\frac{ab}{m\delta_n^{d-2}} + \frac{b^2}{m^2n^2\delta_n^{2d}} + \frac{v}{n}.$$

If we assume that the computational effort is primarily spent in the outer loop in generating samples of  $Y$  (average  $\gamma$  per sample) and in the inner loop in generating samples of  $X$  conditional on  $Y$  (average effort  $\beta$  per sample), and that the effort required to compute  $\bar{X}_m(Y)$  once the associated samples of  $Y$  and  $X$  are known, is negligible, then, the overall computational effort is close to  $n(\gamma + m\beta)$ . We then look for  $m$ ,  $n$  and  $\delta_n$  that minimize the mean square error when the total computational budget  $\Gamma = n(\gamma + m\beta)$  and  $\Gamma \rightarrow \infty$ .

In this formulation, for conducting order of magnitude analysis, analysis simplifies if we set  $\gamma = 0$  and  $\beta = 1$ . (The conclusions do not essentially change by this parameter restriction).

Then, the mean square error (mse) equals

$$a^2\delta^4 + 2\frac{ab}{\Gamma\delta^{d-2}} + \frac{b^2}{\Gamma^2\delta^{2d}} + \frac{vm}{\Gamma}.$$

It then follows that  $m = 1$  is optimal. Furthermore, optimal  $\delta$  is of order

$$\Gamma^{-\frac{1}{d+2}}.$$

The convergence rate of mse to zero then equals

$$\Gamma^{-\min(1, \frac{4}{d+2})}.$$

Therefore, the kernel based estimation is better than pure nested simulation for  $d \leq 3$  and is worse than nested simulation for  $d \geq 5$ .

## 4 CONCLUSION

In this paper we considered the problem of estimating via simulation the expectation of a non-linear function of a conditional expectation. We noted some of the applications of this in pricing and risk management in finance. We considered this problem in two settings: one where pure nested simulation was used and second where this was combined with a popular kernel method. For both the settings we derived the expressions for the bias and the variance using simplified and heuristic arguments. We showed that if the overall computational budget equals  $\Gamma$ , then as  $\Gamma \rightarrow \infty$ , the mse under the first method decays to zero at rate proportional to  $\Gamma^{-2/3}$ . Under the second method, this depends on the dimensionality of the underlying random vector and equals  $\Gamma^{-\min(1, \frac{4}{d+2})}$  when  $d$  denotes this dimension. Therefore, the second method asymptotically outperforms the first for  $d \leq 3$ , while the first is better for  $d \geq 5$ . As mentioned earlier, the analysis performed is under simplified assumptions. A rigorous presentation combined with numerical support for the results will be presented in an expanded version of this paper.

## A: APPENDIX

**Equation (3) equals (6):** Using Taylor's expansion, we have along the set  $\{EX|Y \leq b, \bar{X}_m(Y) \leq b\}$ ,

$$G(\bar{X}_m(Y)) - G(EX|Y) = (\bar{X}_m(Y) - EX|Y)G'(EX|Y) + \frac{(\bar{X}_m(Y) - EX|Y)^2}{2}G''(EX|Y) + \text{smaller order terms.}$$

Then,  $E(G(\bar{X}_m(Y)) - G(EX|Y))I(EX|Y \leq b, \bar{X}_m(Y) \leq b)$  equals

$$= E(\bar{X}_m(Y) - EX|Y)G'(EX|Y)I(EX|Y \leq b, \bar{X}_m(Y) \leq b) \quad (14)$$

$$+ \frac{1}{2}E(\bar{X}_m(Y) - EX|Y)^2G''(EX|Y)I(EX|Y \leq b, \bar{X}_m(Y) \leq b) \quad (15)$$

$$+ \text{smaller order terms.} \quad (16)$$

Note that

$$E(\bar{X}_m(Y) - EX|Y)G'(EX|Y)I(EX|Y \leq b) = 0.$$

Hence, (14) equals

$$-E(\bar{X}_m(Y) - EX|Y)G'(EX|Y)I(EX|Y \leq b, \bar{X}_m(Y) > b).$$

Recall that  $\bar{X}_m(Y)$  can be expressed as  $EX|Y + \frac{W}{\sqrt{m}}$  where conditional on  $Y$ ,  $W$  has a Gaussian distribution with mean zero and variance  $\sigma^2(Y)$ . The above equation may then be re-expressed as

$$-E\left(\frac{W}{\sqrt{m}}\right)G'(EX|Y)I(EX|Y \leq b, EX|Y > b - \frac{W}{\sqrt{m}}). \quad (17)$$

Let  $f_{H,W}(\cdot, \cdot)$  denote the joint pdf of  $H = EX|Y$  and  $W$ . Then, (17) equals

$$-\int_0^\infty \frac{w}{\sqrt{m}} \int_{b-w/\sqrt{m}}^b G'(h) f_{H,W}(h, w) dh dw. \quad (18)$$

By using the Taylor series expansion

$$f_{H,W}(h, w) = f_{H,W}(b, w) + (h - b) \frac{\partial f_{H,W}(b, w)}{\partial h} + \text{smaller order terms} \quad (19)$$

(assuming that this density expansion exists) in (18), we see that under mild conditions, it equals

$$-\frac{G'(b)f_H(b)}{m}E(W^2I(W \geq 0)|H = b) + \text{smaller order terms.}$$

Note that

$$E(W^2I(W \geq 0)|H = b) = E(E(W^2I(W \geq 0)|Y)|H = b) = \frac{1}{2}E(\sigma^2(Y)|H = b).$$

Hence, (14) equals

$$-\frac{1}{2} \frac{G'(b)f_H(b)}{m} E(\sigma^2(Y)|H = b) + \text{smaller order terms.} \quad (20)$$

Now consider (15). It may be re-expressed as the difference of

$$\frac{1}{2}E(\bar{X}_m(Y) - EX|Y)^2G''(EX|Y)I(EX|Y \leq b) \quad (21)$$

and

$$\frac{1}{2}E(\bar{X}_m(Y) - EX|Y)^2G''(EX|Y)I(EX|Y \leq b, \bar{X}_m(Y) > b). \quad (22)$$

Conducting the analysis as for evaluating (14), it can be seen that (22) is of smaller order compared to  $1/m$ . Since, (21) equals

$$\frac{1}{2m} E \sigma^2(Y) G''(EX|Y) I(EX|Y \leq b),$$

(6) follows from (20) and (15).

**Equation (4) equals (7):** Equation (4) equals

$$-E(G(EX|Y)I(EX|Y \leq b, \bar{X}_m(Y) \geq b)),$$

and may be re-expressed as

$$- G(b)P(EX|Y \leq b, EX|Y \geq b - \frac{W}{\sqrt{m}}) \quad (23)$$

$$- G'(b)E(EX|Y - b)I(EX|Y \leq b, EX|Y \geq b - \frac{W}{\sqrt{m}}), \quad (24)$$

plus smaller order terms. We may re-express (23) as

$$-G(b) \int_0^\infty \int_{b-w/\sqrt{m}}^b f_{H,W}(h,w) dh dw \quad (25)$$

Again, by using the Taylor series expansion as in (19), in (25), we see that it equals

$$-\frac{G(b)f_H(b)}{\sqrt{m}} E(WI(W \geq 0)|H = b) + \frac{G(b)}{2m} \int_0^\infty w^2 \frac{\partial f_{H,W}(b,w)}{\partial b} + \text{smaller order terms.}$$

By interchanging the partial derivative and the integral, the second term above may be re-expressed as

$$\frac{G(b)}{2m} \frac{d}{db} f_H(b) E(W^2 I(W \geq 0)|H = b). \quad (26)$$

As before,

$$E(W^2 I(W \geq 0)|H = b) = \frac{1}{2} E(\sigma^2(Y)|H = b).$$

Therefore, (25) can be re-expressed as

$$\frac{1}{4m} G(b) \frac{d}{db} f_H(b) E(\sigma^2(Y)|H = b) \quad (27)$$

plus smaller order terms.

Now we evaluate the dominant terms in (24). It may be re-expressed as

$$-G'(b) \int_0^\infty \int_{b-w/\sqrt{m}}^b (h-b) f_{H,W}(h,w) dh dw.$$

By considering the Taylor series expansion of  $f_{H,W}(h,w)$  this can be seen to be

$$\begin{aligned} &\approx G'(b) \int_0^\infty \frac{w^2}{2m} f_{H,W}(b,w) dw, \\ &= \frac{G'(b)f_H(b)}{2m} E(W^2 I(W \geq 0) | H = b) \\ &= \frac{G'(b)f_H(b)}{4m} E(\sigma^2(Y) | H = b). \end{aligned}$$

**Equation (5) equals (8):** Along the set  $\{EX|Y > b, \bar{X}_m(Y) \leq b\}$ , we have

$$G(\bar{X}_m(Y)) = G(b) + (\bar{X}_m(Y) - b)G'(b) + \frac{(\bar{X}_m(Y) - b)^2}{2} G''(b) + \text{smaller order terms.}$$

It follows that (5) equals

$$G(b)P(EX|Y > b, EX|Y + W/\sqrt{m} \leq b) \quad (28)$$

$$+ G'(b)E(W/\sqrt{m} + EX|Y - b)I(EX|Y > b, EX|Y + W/\sqrt{m} \leq b) \quad (29)$$

+ a smaller order terms. Equation (28) in turn equals,

$$G(b) \int_{-\infty}^0 \int_b^{b-w/\sqrt{m}} f_{H,W}(h,w) dh dw \quad (30)$$

As before, by using the Taylor series expansion of  $f_{H,W}(h,w)$  at  $f_{H,W}(b,w)$ , in (30), we see that it equals

$$- \frac{G(b)f_H(b)}{\sqrt{m}} E(WI(W \leq 0) | H = b) + \frac{G(b)}{2m} \int_{-\infty}^0 w^2 \frac{\partial f_{H,W}(b,w)}{\partial b} + \text{smaller order terms.}$$

By interchanging the partial derivative and the integral, the second term above may be re-expressed as

$$\frac{G(b)}{2m} \frac{df_H(b)}{db} E(W^2 I(W \leq 0) | H = b). \quad (31)$$

Note that

$$E(W^2 I(W \leq 0) | H = b) = E(E(W^2 I(W \leq 0) | Y) | H = b) = \frac{1}{2} E(\sigma^2(Y) | H = b).$$

Therefore, (31) can be re-expressed as

$$\frac{G(b)}{4m} \frac{df_H(b)}{db} E(\sigma^2(Y) | H = b). \quad (32)$$

Now we evaluate the dominant terms in (29). It may be re-expressed as

$$G'(b) \int_{-\infty}^0 \int_b^{b-w/\sqrt{m}} (w/\sqrt{m} + h - b) f_{H,W}(h,w) dh dw.$$

By considering the Taylor series expansion of  $f_{H,W}(h,w)$  at  $f_{H,W}(b,w)$ , this can be seen to be

$$\begin{aligned}
 &\approx G'(b) \int_{-\infty}^0 f_{H,W}(b,w) \left( \int_b^{b-w/\sqrt{m}} (w/\sqrt{m} + h - b) dh \right) dw, \\
 &= -G'(b) \int_{-\infty}^0 \frac{w^2}{2m} f_{H,W}(b,w) dw, \\
 &= -\frac{G'(b)f_H(b)}{2m} E(W^2 I(W \leq 0) | H = b) \\
 &= -\frac{G'(b)f_H(b)}{4m} E(\sigma^2(Y) | H = b).
 \end{aligned}$$

## REFERENCES

- Bosq, D. 1998. *Nonparametric statistics for stochastic processes*. Springer, New York.
- Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*. Springer.
- Gordy, M., and S. Juneja. 2006. Efficient Simulation for Risk Measurement in Portfolio of CDOs. *In Proceedings of the 2006 Winter Simulation Conference*, eds. L. R. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. Fujimoto, 1989-1996. Piscataway, New Jersey: IEEE Inc. 749-756.
- Gordy, M., and S. Juneja. 2008. Nested simulation in portfolio risk measurement. Finance and Economics Discussion Series, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington, D.C., 2008-21.
- Lee, S. H. 1998. *Monte Carlo expectation of conditional expectation quantiles*. Ph.D. thesis, Department of Operations Research, Stanford University, Palo Alto, California.
- Li, Q., and J. S. Racine. 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, New Jersey.
- Pagan, A., and A. Ullah. 1999. *Nonparametric Econometrics*. Cambridge, UK

## AUTHOR BIOGRAPHIES

**L. JEFF HONG** is associate professor in the Department of Industrial Engineering and Logistics Management at the Hong Kong University of Science and Technology (HKUST). He received a Ph.D. in industrial engineering and management sciences from Northwestern University in 2004. His research interests include stochastic simulation, financial risk management, stochastic optimization, and operations and logistics management. He is currently an associate editor for *Operations Research*, *Naval Research Logistics* and *ACM Transactions on Modeling and Computer Simulation*. His email address is <hongl@ust.hk>.

**SANDEEP JUNEJA** is associate professor in the School of Technology and Computer Science at Tata Institute of Fundamental Research. He received his Ph.D. in operations research from Stanford University. His research interests include applied probability, computational finance, stochastic simulation and rare-event analysis. He serves as an associate editor for *Mathematics of Operations Research* and *ACM Transactions on Modeling and Computer Simulation*. His email address is <juneja@tifr.res.in>.